Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
International Journal of Forecasting 24 (2008) 659 – 678
international journal of forecasting
www.elsevier.com/locate/ijforecast

# A nonlinear mixed effects model for the prediction of natural gas consumption by individual customers

Marek Brabec *, Ondřej Konár, Emil Pelikán, Marek Malý

*Institute of Computer Science, Academy of Sciences of the Czech Republic Pod Vodárenskou věží 2, 182 07, Praha 8, Czech Republic*

**Abstract**

This study deals with the description and prediction of the daily consumption of natural gas at the level of individual customers. Unlike traditional group averaging approaches, we are faced with the irregularities of individual consumption series posed by inter-individual heterogeneity, including zeros, missing data, and abrupt consumption pattern changes. Our model is of the nonlinear regression type, with individual customer-specific parameters that, nevertheless, have a common distribution corresponding to the nonlinear mixed effects model framework. It is advantageous to build the model conditionally. The first condition, whether a particular customer has consumed or not, is modeled as a consumption status in an individual fashion. The prediction performance of the proposed model is demonstrated using a real dataset of 62 individual customers, and compared with two more traditional approaches: ARIMAX and ARX.


*Keywords:* Individual gas consumption; Nonlinear mixed effects model; ARIMAX; ARX; Generalized linear mixed model; Conditional modeling

## 1. Introduction

Deregulation of the natural gas market leads to greater freedom for its participants, but also increases the risk of financial losses. Therefore, all subjects, whether they be shippers, traders, or consumers, would like to minimize this risk. One of the paths to risk reduction involves a good consumption estimate.

In this paper, we focus on gas consumption modeling and forecasting for individual customers. Why model these data individually? Sums or averages of large groups of customers are commonly modeled and predicted (Schwarz, Kubessa, & Fuhrberg-Baumann, 2002; Bailey, 2000; Gerbec, Gašperič, Šmon, & Gubina, 2004). In this context, various sophisticated approaches to hourly resolution data have been applied, see e.g. Soares and Medeiros (2008-this issue), who deal with electricity load modeling. However, there are situations for which an individual approach is important; for instance, when an advance payment structure must be constructed or when the temperature response variability among individual customers in a trader's portfolio is investigated, both for marketing purposes and for pipeline capacity evaluations. Individual modeling and predictions are very relevant, especially during periods of price changes and in market deregulation scenarios, which

* Corresponding author. Tel.: +420 266 053 710.
*E-mail address:* mbrabec@cs.cas.cz (M. Brabec).

0169-2070/$ - see front matter © 2008 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2008.08.005

are currently of particular interest in many countries. In addition, there are many situations for which one is only interested in the total or average consumption, but is driven to individual modeling and prediction for various technical reasons. Let us examine one particular example. Gas traders might easily acquire time-varying portfolios of customers and need to predict total consumption, e.g. to secure adequate gas volume in the future. Time changes in the portfolio make the meaning of the total consumption time series somewhat unclear and not particularly well defined. This fact has two types of consequences:

- For model training; these are related to changes in the composition of the total sum in the historical data used for model training.
- For prediction; these occur when the total predicted value consists of different individual consumptions.

In a real situation, both problems are present, calling for a "bottom-up" approach based on individual modeling, followed by summing the estimates or predictions. In this way, individual modeling might be viewed as a means of obtaining a proper assessment of the total, and hence it might occasionally be very relevant, even if one is interested in modeling only the sum across customers. The aforementioned scenarios are stressed even more by the technical issue of missing data. Missing individual consumption data leads again to a time-varying composition of the total, and problems similar to those mentioned above. If there are only a few missing values and/or the mechanism of "missingness" is simple enough (e.g. MCAR, see Little & Rubin, 1987), one might ignore the problem. If this is not the case and one does not want to go into the "missingness" modeling directly, an individual approach might be necessary.

In addition, individual modeling provides a lot more information than modeling the total consumption, including information of direct practical interest, e.g. about inter-individual variability, as well as information which is useful for model checking and criticism. Furthermore, it is clear that once individual estimates and/or predictions are available, one can form various aggregates of direct practical interest, for instance, sums across groups of customers, sums across time — constructing e.g. monthly or various irregular time interval estimates — or sums across both time and customers. Of course, when the desired aggregation is known in advance, one can focus on modeling only the aggregated data. However, when it is not known and/or there are several overlapping aggregations that need to be treated in a consistent way, one would prefer to model the individual data and aggregate the results. In general, modeling overlapping aggregations separately does not prevent inconsistencies and conflicts. All of these considerations lead us to prefer individual modeling for certain particular cases. In the vast majority of published approaches, forecasting is performed for various groups of customers such as residential, commercial, and industrial. Many approaches to modeling on the group level have been described and applied to both gas and electricity consumption data. These include regression and econometric models of varying complexity (Bartels, Fiebig, & Nahm, 1996; Gil & Deferrari, 2004), the classical time series approach and its modifications (Sánchez-Úbeda & Berzosa, 2007), the Bayesian approach (Bauwens, Fiebig, & Steel, 1994), and artificial neural network models (Suykens et al., 1996; Gerbec et al., 2004; da Silva, Ferreira, & Velasquez, 2008-this issue). The inclusion of ARMA terms in regression models enables one to capture the effects of unmeasurable variables, and often improves model accuracy (Wang, 2004). Similar tasks are solved in the context of modeling and forecasting electricity load demands. For instance, Liu, Chen, Liu, Liu and Harris (2006) explore a semi-parametric time series approach, Amaral, Souza and Stevenson (2008-this issue) deal with nonlinear time series models, Taylor (2008-this issue) considers exponential smoothing and ARIMA modelling, Cottet and Smith (2003) build Bayesian models and forecasts, Harvey and Koopman (1993) make use of time-varying splines, and Dordonnat, Koopman, Ooms, Dessertaine and Collet (2008-this issue) explore another quite general state space approach.

Although similar modeling methods can be used to build predictions for individual customers, the models are more complex in order to capture both customer-specific and common group variability. To the best of our knowledge, information on such models for individual customers is very limited in the literature.

A second interesting problem besides individual modeling emerges from the methods used to measure consumption. For instance, the customers in Slovakia are divided into three classes, according to the type of consumption measurement. The first class is composed mainly of large commercial and industrial customers whose consumptions are measured at a daily resolution, the data for which are available on-line. Measurements for the second class, which also consists of large commercial entities, are also obtained on a daily basis, but the data are read manually once a month using an off-line regime. The last class consists of small commercial and residential customers whose meter readings are typically taken at long intervals, usually annually.

In the present paper, we focus on customers from the second class. Although the daily consumption is always known at the end of each month, for various practical purposes (e.g. economical and technological balancing, nomination processes etc.) the company requires an estimation of individual daily consumption during the period from the last reading to a given day. The estimates have to be evaluated only at a time when all explanatory variables, e.g. temperature, are available.

This generates an interesting and not completely standard forecasting problem. The forecast horizon varies from 1 to 31 days, and the model can be updated sequentially every month with the arrival of a new monthly batch of data. Useful exogenous variables are available on-line for each day of the month, so that they need not be predicted.

Since the purpose of gas usage (cooking, water heating, space heating, technologies) varies among customers, there is substantial inter-individual variability in gas consumption profiles across various timescales, e.g. weekly, monthly, and annually. Therefore, outdoor temperature and calendar effects seem to be the most important available predictors. For calendar effects, we used only regular ones, namely the day of the week effect. In principle, one might think of improving the model performance by including the effects of bank holidays, as well as the Christmas and Easter vacation periods. In order to keep the model simple while minimizing the loss of overall performance, we did not do this. Moreover, the data available are not detailed enough to comfortably estimate irregular effects. For instance, for the Christmas effect we would have access to only one season in the historical data to train the model. In addition, we did not use any other external macroeconomic variables such as gas prices, inflation, money exchange rates, GDP, or unemployment rates. This was intentional. We wanted to concentrate on a rather general model of consumption relative to temperature and calendar effect relationships. Macroeconomic effects modeling is interesting and important, but it requires a long history of data, longer than we have at the moment. It also requires certain additional inputs, e.g. careful inflation corrections, as well as at least rough a priori ideas about the shape of the impulse response function (more than are currently available).

In the present study, we used available data measured on a daily basis from 62 individual commercial customers. Their consumption was monitored continuously from January 1, 2005 to June 29, 2007 as part of a large scale load profile construction project organized by SPP (contract No. 421, ICS 2005–2007). We used Slovak daily average temperatures. The goal was to build a general model of consumption and compare its performance with other, more standard models. To this end, we used 62 available individual trajectories. Our primary goal in this paper was to estimate individual consumption, *not* the consumption of the whole customer segment. If such a model were implemented in practice, it would be used for as many customers as is needed for a particular segment, typically dozens to hundreds.

For the purpose of this study, we split the data into an *historical* dataset, the first 18 months, and an *evaluation* dataset, consisting of twelve parts corresponding to the remaining twelve months. To conform to real life situations, model parameters were reestimated every month during the evaluation period. The training set, i.e. the part of the data used for parameter estimation, consisted of the historical dataset, and was expanded every month concomitantly with the sequential monthly arrival of new data, as is outlined in Fig. 1. After every reestimation, the month following the current training set serves as the data source for the evaluation.

There are many nonstandard features in the data. These are illustrated in Fig. 2, which shows the overall situation for the historical data as an overlay of the individual consumption time series. From this, one can get a general feeling for the high inter-individual variability.

In Fig. 3, we further show four typical consumption trajectories over the whole historical period. From this, we can see an obvious annual periodicity in the data. Fig. 4 shows one of these trajectories in detail over a shorter period, suggesting some weekly repeating patterns.

The patterns are imposed by human activity and weather patterns. Clearly, weekly periodicity is to be expected a priori if one realizes that we are dealing with commercial customers who demonstrate many

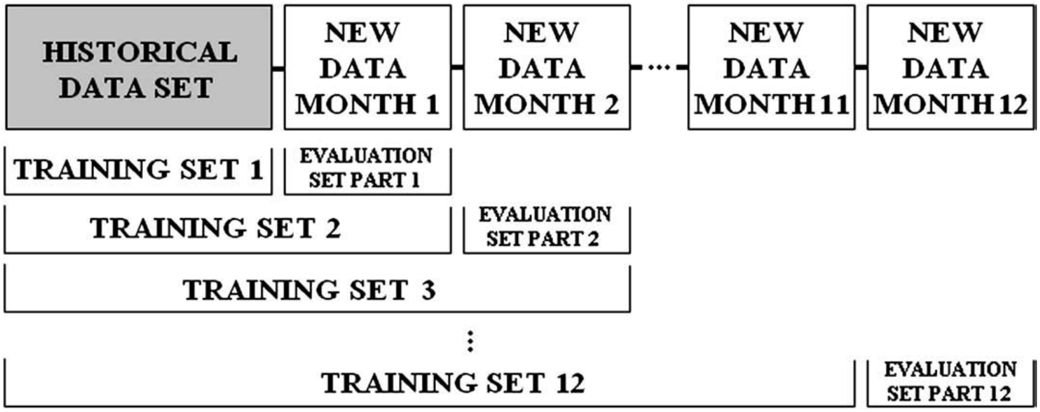


Fig. 1. Diagram of the monthly expansion of the training dataset.

changes in their economic and other activities across weekdays. Due to this fact, consumption varies by several orders of magnitude. Nevertheless, the periodicity is far from completely regular. We can also clearly see substantial inter-individual heterogeneity in the consumption profiles. Not only are the levels substantially different, but in addition, the shapes are not the same; in other words, the interaction of time or temperature, which is related to time and individual behavior, is substantial. Moreover, there are non-negligible consumption pattern changes even for a given individual. These are connected to many processes that the gas companies have no information about — for instance, technological changes in small commercial companies, different amounts of individual economic activity depending on the available contracts and their processing, etc. There are also occurrences of zero consumption related to maintenance, individual company holidays, economic breakdowns, individual company difficulties, etc. The proportion of zero consumption days varies considerably between different customers (Fig. 5).

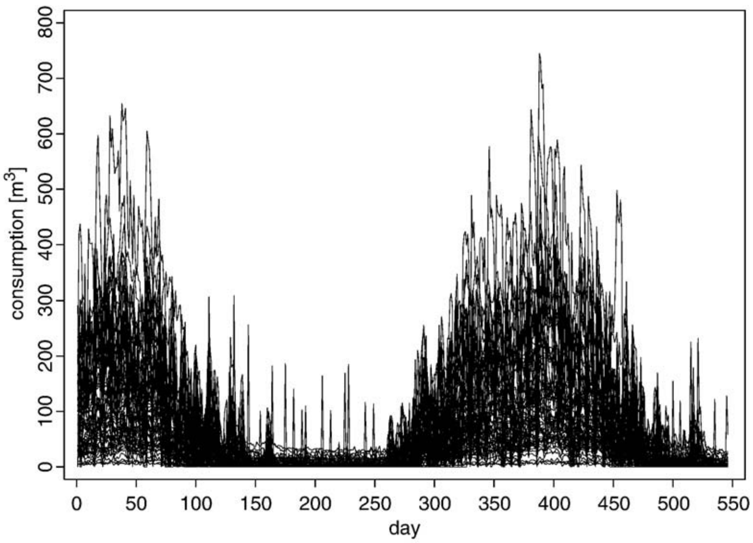


Fig. 2. Empirical behavior of the individual historical consumption data from January 1, 2005 ($Day = 1$) to June 30, 2006 ($Day = 545$); overlay of 62 individual consumption trajectories.

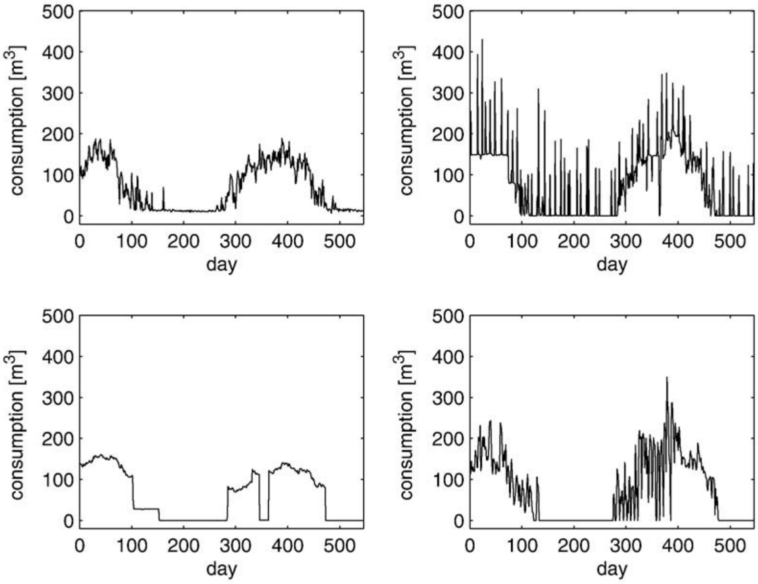


Fig. 3. Example of four individual historical consumption trajectories from January 1, 2005 (*Day*=1) to June 30, 2006 (*Day*=545).

For various technical reasons related particularly to the measurement process, there are also missing data in both the historical and evaluation portions (up to 10% per individual).

The rest of this paper is organized as follows. In Section 2.1 we will postulate a nonlinear statistical model with random parameters addressing the problems mentioned above.

Sections 2.2.1 and 2.2.2 describe two classical time series (TS) approaches (ARIMAX and ARX) that are used here for a performance comparison with the developed model. The motivation for this selection is

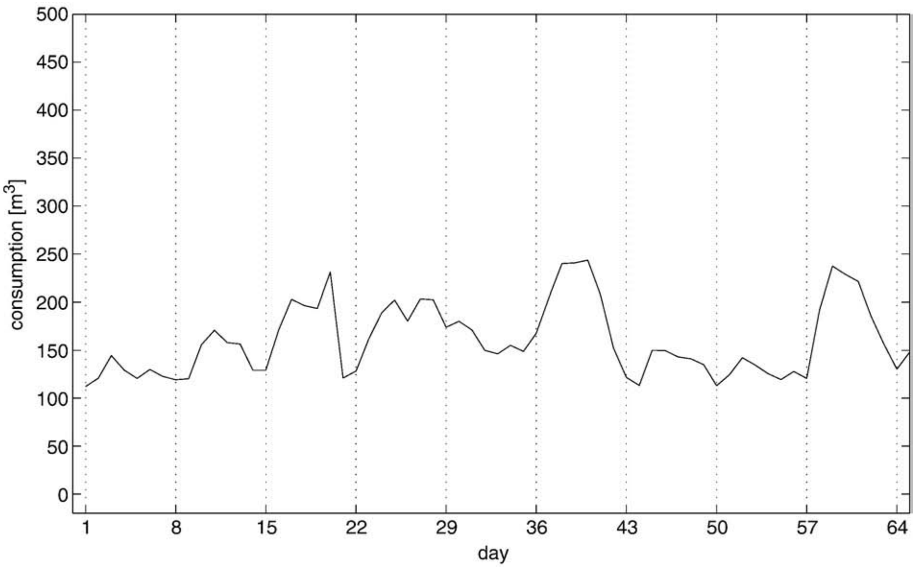


Fig. 4. “Zoomed” example of an individual historical consumption trajectory from January 1, 2005 (*Day*=1) to March 5, 2005 (*Day*=64).

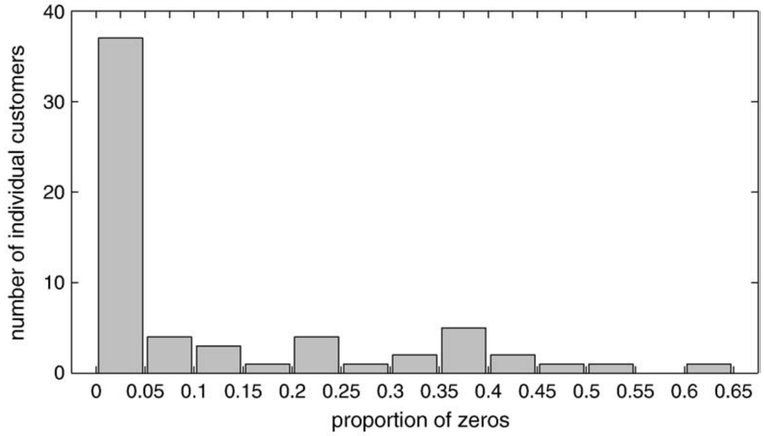


Fig. 5. Distribution of customers according to the proportion of days with zero consumption in the historical data.

discussed in Section 2.2.3. The prediction performance and a comparison of all approaches are presented in Section 3. In Section 4, the results are summarized and possible further developments are discussed.

## 2. Models

### 2.1. Nonlinear mixed effects model, NLME

In this section, we describe a parametric statistical model with several structural parts built in such a way that each of them have a real world interpretation. The fact that the model is composed of several interpretable parts actually causes its modularity. This is advantageous in model building, as well as in connection with its "serviceability". For instance, one can implement different versions of the model at different geographic locations with different temperature response functions suitable to the particular consumption pattern and local situation. The structured nature of the model is also useful for model checking and criticism. Furthermore, on various occasions, one can gain a lot of qualitative and quantitative insight from looking at individual parameters.

Our model is specified conditionally. It therefore consists of two parts:

- a conditional model of consumption, given that it is positive $(Y|Y>0)$, see Section 2.1.1, model (1).
- a marginal probability of zero consumption $P(Y=0)$, see Section 2.1.2, model (2).

Both parts are individual-specific and amenable to prediction for future days once the model parameters have been estimated using historical data. They can be combined to obtain an estimate for a particular day. In fact, the two parts model integrates two aspects of a mixture with a degenerated second component: consumption with a probability of $1–P(Y=0)$ and no consumption with a probability of $P(Y=0)$. In other words, the consumption is effectively modeled as a mixture $(Y|Y>0).P(Y>0)+0.P(Y=0)$.

#### 2.1.1. Conditional model of consumption

Our conditional model of the consumption $Y_{it}$ of the $i$-th customer ($i=1,\ldots n$) on day $t$ corresponding to a weekday $k$, where $t=1,\ldots, T$, given that $Y_{it}>0$, i.e., given that the particular customer consumes some gas on day $t$, is specified through the individual-specific vector of parameters $\theta_i=\{\alpha_i, \delta_{i0}, \delta_{i1}, \delta_{i2}, \delta_{i3}, \delta_{i4}, \delta_{i5}, \delta_{i6}, \lambda_i, \mu_i, \omega_i, \gamma\}$ as follows:

$$\begin{aligned}(Y_{it}|Y_{it}>0) = g(t;\theta_i)exp(\varepsilon_{it}) = [&exp(\alpha_i+\delta_{ik}) \\ &+exp(\lambda_i)G(T_t+\gamma T_{t-1};\mu_i,\omega_i)]exp(\varepsilon_{it})\end{aligned} \tag{1}$$

where the days of the week are indexed by $k=0,1,\ldots,6$, with Sunday corresponding to $k=0$. The average temperature for day $t$ is denoted by $T_t$. For identifiability, we have similarly restricted the $\delta_{ik}$s to general linear models, e.g. in the context of ANOVA models (Searle, 1971). We used the sum restriction $\sum_{k=0}^{6}\delta_{ik}=0$, implying $\delta_{i6}=-\sum_{k=0}^{5}\delta_{ik}$.

This model allows for an individual-specific range of temperature responses. This feature is implemented by two model components. In fact, model (1) has both an individual-specific median response minimum given by $\exp(\alpha_i+\delta_{ik})$ and a maximum temperature-related response range given by $\exp(\lambda_i)G(T_t+\gamma\cdot T_{t-1};\mu_i,\ \omega_i)$. This is dictated by the large heterogeneity in minima and temperature response sizes, which is due to the high variability among customers illustrated in Fig. 2. The consumption minimum also changes over time. Its median level $\exp(\alpha_i)$ is modified multiplicatively to capture the different working schedules of different customers depending on the weekday. Some of the commercial customers work throughout the week, while others work only on working days. This results in very different weekly profiles. Estimates of the median week profiles can be obtained very easily simply by exponentiating the model coefficients $\delta_{ik}$. The resulting $\exp(\delta_{ik})$ represents a part of the profile corresponding to customer $i$ and the $k$-th day of the week. Individual week profiles are plotted in Fig. 6.

The shape of the temperature response is given by the function $G(x;\ \mu_i,\ \omega_i)$. Clearly, consumption tends to decrease with increasing temperature such that $G$ should generally be nonincreasing with respect to $x$. To reflect physically plausible behavior, it should have lower and upper asymptotes. The role of nonlinearity in temperature response has been discussed by Cancelo, Espasa and Grafe (2008-this issue). For identifiability, this should be scaled appropriately. With $\min_x G(x;\ \mu_i,\ \omega_i)=0$ and $\max_x G(x;\ \mu_i,\ \omega_i)=1$, we attain a good physical interpretation of the model structure. In principle, we might use any cumulative distribution function $F(x;\ \mu_i,\ \omega_i)$, take its complement, and use the resultant value as $G(x;\ \mu_i,\ \omega_i)=1-F(x;\ \mu_i,\ \omega_i)$. Location-scale families are particularly convenient and appealing. The temperature response's derivative $\frac{\partial}{\partial x}G(x;\mu_i,\omega_i)$ is asymmetric around the location of its peak, with a faster rate of decrease than increase, as can be seen in Fig. 7. Therefore we have chosen the smallest extreme value distribution (SEV), i.e. the Gumbel distribution for $F$. In other situations, e.g., for geographic locations with a much milder climate, other $F$s might be more suitable. In other words, we used $G(x;\mu_i,\omega_i) = exp\left(-exp\left(\frac{x-\mu_i}{\omega_i}\right)\right)$. We have tested other models in the course of model searching, and found SEV to be the best among those tried. A nonparametric description of the consumption–temperature relationship would also be possible, and would be appealing when one has a longer history of data available. In actual fact, the whole model would become semiparametric, and would be much more complicated to estimate. It would also introduce the practical disadvantage of not being able to extrapolate beyond the minimum and maximum temperature points observed for the training data.

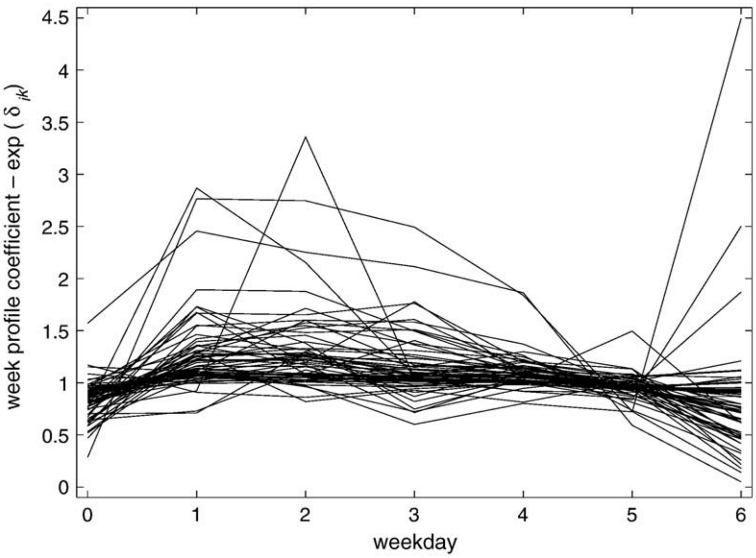


Fig. 6. Week profiles, $\exp(\delta_{ik})$, $i=1\ldots n$, $k=0\ldots6$, estimated from model (1) using historical data ($k=0$ corresponds to Sunday).

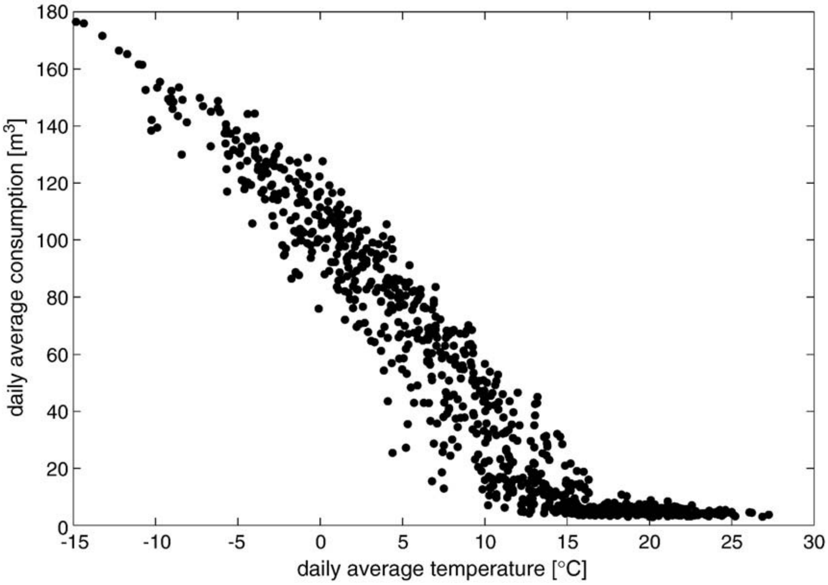


Fig. 7. Mean temperature response by averaging empirically across customers using historical data.

On the basis of both common-sense considerations and some formal testing results obtained from the historical data, we prefer to use not only the current daily average temperature $T_t$, but also its one-day lag in the form of the weighted average $T_t+\gamma\cdot T_{t-1}$. To ease the computational burden connected with model estimation, the weight $\gamma$ is not individual-specific, unlike the other components of model (1).

We assumed a normal distribution for individual-specific parameters: $\alpha_i \sim N(A,\ \sigma_A^2)$; $\delta_{ik} \sim N(D_k,\ \sigma_{D_k^2})$, $k=0,1,\ldots 6$; $\lambda_i \sim N(L,\sigma_L^2)$; $\mu_i \sim N(M,\ \sigma_M^2)$; and $\omega_i \sim N(U,\ \sigma_U^2)$; and also for the model error $\varepsilon_{it} \sim N(0,\ \sigma^2)$. Furthermore, we assumed that all of them were independent of one another. Clearly, this specification ensures that the consumption estimates from the model remain positive, as they should. Within-individual distributions that are conditional on $\theta_i$ are lognormal. A skewed distribution corresponds to the empirical consumption behavior, and, on this account, is in agreement with other previously published gas consumption statistical models (Vondráček et al., 2008). The assumed independence might be intuitively appealing in the sense that a model with a diagonal random effects covariance matrix will tend to be more flexible, i.e. less dependent on the relationships among the parameters. However, we might ask whether it is plausible to assume the independence and normality of the random parameters in the case of this particular data. To this end, we might check bivariate plots and/or correlations among different random parameters, and examine their histograms. There are many of these, since the model is reestimated each time a new monthly batch of data arrives. Fig. 8 illustrates one particular case, which shows no obvious correlation and no substantial departures from normality.

Overall, we are dealing with a *nonlinear mixed effects model* (NLME) here (Pinheiro & Bates, 2000), with the obvious reparametrization $\rho_i=R+r_i$. The generic individual-specific or random parameter $\rho_i$ is then expressed with the fixed effect $R$ as its population mean and the random effect $r_i$ as its departure from the mean. Clearly, the parameters of the conditional model (1) can be estimated quite easily by fitting the model to historical nonzero data. We used the nlme library, available in both S-PLUS (Anonymous, 2001) and R (R Development Core Team, 2008), for related computations. In other words, we estimated model parameters via maximum likelihood. Missing values were excluded from the training process.

### 2.1.2. *A model for marginal probability of zero consumption*

The marginal probability of a zero consumption day is modeled in both a time-varying and an individual-

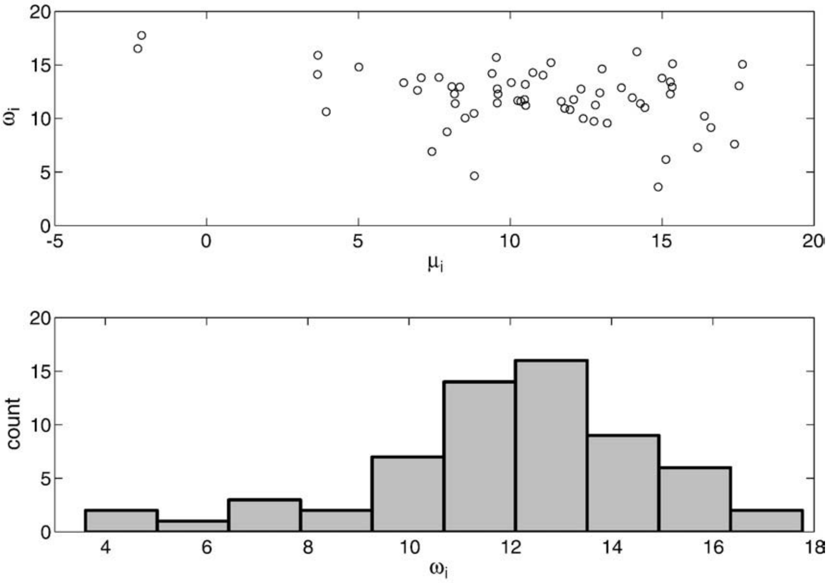


Fig. 8. Bivariate plot of $\omega_i$ versus $\mu_i$ (upper panel) and a histogram of the distribution of $\omega_i$ (lower panel). Both are for estimates obtained from the initial training dataset.

specific manner. In fact, we allowed it to depend on both temperature and the weekday, to reflect different tendencies to switch off during different times of the year and/or during the week. We used logistic regression with the zero consumption day indicator as the dependent variable. Using $\pi_{it}=P(Y_{it}=0)$ for brevity, we formulate the marginal model as follows:

$$log\left(\frac{\pi_{it}}{1-\pi_{it}}\right) = \beta_{i0} + \beta_{i1}\cdot T_t + \textstyle\sum_{k=0}^{6} \beta_{2+k}\cdot I(\text{day } t \text{ is of type } k)$$
$$I(Y_{it}=0) \sim Bin(1,\pi_{it}), \tag{2}$$

with the sum restriction on the weekday portion for identifiability, as in Section 2.1.1. Throughout the paper, we use $I(.)$ to denote an indicator function having a value of 1 when the condition listed in the argument is satisfied and 0 when it is otherwise, and we denote the natural, i.e. $e$-based logarithm, by log.

Both intercepts and slopes for the temperature are random and independent, $\beta_{i0}\sim N(B_0,\ \sigma_0^2)$, $\beta_{i1}\sim N(B_1,\ \sigma_1^2)$. Weekday-specific corrections do not have random components. We tried an expanded model by adding random weekday effects, but it demonstrated worse fitting of the data (in terms of AIC, e.g. for the historical dataset). The distribution of the indicator that $Y_{it}=0$, given all fixed and random parameters, is Bernoulli with a probability $\pi_{it}$. Its logit was modeled as prescribed in Eq. (2). Individual-specific random effects were used to allow for different zero consumption tendencies among individuals. Fig. 5 gives the impression of heterogeneity, which is substantial when comparing the differences between overall levels. The temperature dependence is largely different for different individuals. In other words, there is substantial interaction of individual and temperature effects. See Fig. 9 for logits of zero consumption day probabilities. The individual tracks are substantially different, and not only with regard to levels; they are not even parallel. In fact, there are a group of individuals that are not far from being parallel, but there are many substantial exceptions. All of this suggests that the interaction of time with individual effects is very important.

Model (2) is a logistic regression with independent random effects, as an instance of relatively simple generalized linear mixed effects models (GLMM, McCulloch & Searle, 2001). It can be fitted via the Penalized Quasi-Likelihood (PQL) approximate

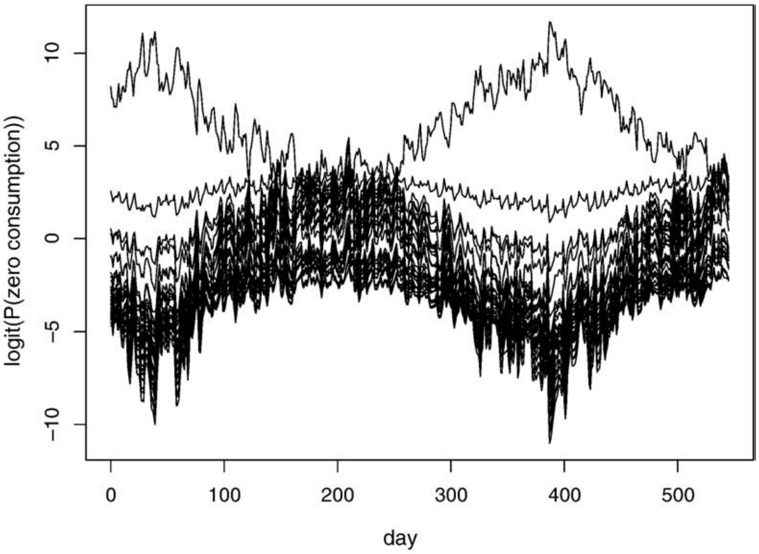


Fig. 9. Logits of $\pi_{it}$'s individual trajectories estimated from historical data.

approach (Breslow & Clayton, 1993; Wolfinger & O'Connell, 1993). We used Ripley's S-PLUS/R function glmmPQL (Venables & Ripley, 1999) for practical computations.

### 2.1.3. Combining conditional and marginal parts

Consumption can be estimated as either

- $\hat{Y}_{it,S}=(1\text{-}\hat{\pi}_{it})\cdot g(t;\ \hat{\theta}_i)$, which we will call the "smooth type"

or

- $\hat{Y}_{it,C}=(1-I(\hat{\pi}_{it}>\text{c}))\cdot g(t;\ \hat{\theta}_i)$ for some cutoff $c$, which we will call the "cutoff type".

Under the lognormal assumptions for a given $\theta_i$ that are part of (1), both of these approaches provide estimates of the median, not the mean. The mean can be advocated only as an approximation, e.g., the Laplace approximation (Ghosh, Delampady, & Samanta, 2006). In fact, the expected value $E(Y_{it}|Y_{it}>0,\ \theta_i)$ would be estimated by $g\left(t;\hat{\theta}_i\right)\cdot exp\left(\frac{\sigma^2}{2}\right)$. In principle, one might try to incorporate the inherent correction into the prediction, although we do not recommend this — at least not by directly exchanging $\hat{\sigma}^2$ for the unknown $\sigma^2$. This is because $\hat{\sigma}^2$ might overestimate the true residual variance $\sigma^2$ due to possible model deficiencies. This would lead to an overestimation of the consumption. Indeed, we attempted this route for our prediction and obtained worse results than without the correction. In the future, it might be interesting to consider more sophisticated $\hat{\sigma}^2$ estimates for improving the consumption estimates, i.e., performance in terms of model fitting to historical data; however, we remain unconvinced that it would improve the predictions substantially.

The first type of estimate/prediction construction might be appealing for smooth behavior, since zero would never be predicted. The second type might be better for individual consumption tracking when the trajectory contains a lot of zeros. According to our experience, the cutoff type seems to be slightly better in terms of overall performance, although the overall pattern, as opposed to the individual performance of individuals whose behavior is highly seasonal, is not decisive. Section 3 and Fig. 10 support this idea.

Nevertheless, the difference in the average level is not large, and the cutoff type requires the specification of $c$, as an additional tuning parameter. We performed the selection by trial and error based on the historical dataset, and observed that prediction performance is better when $c>0.5$ than when $c=0.5$, and that changes are minimal until relatively extreme values, say 0.9 or greater, are chosen. Model prediction abilities mostly deteriorated for the summer months, with low consumption, and marginally improved during months

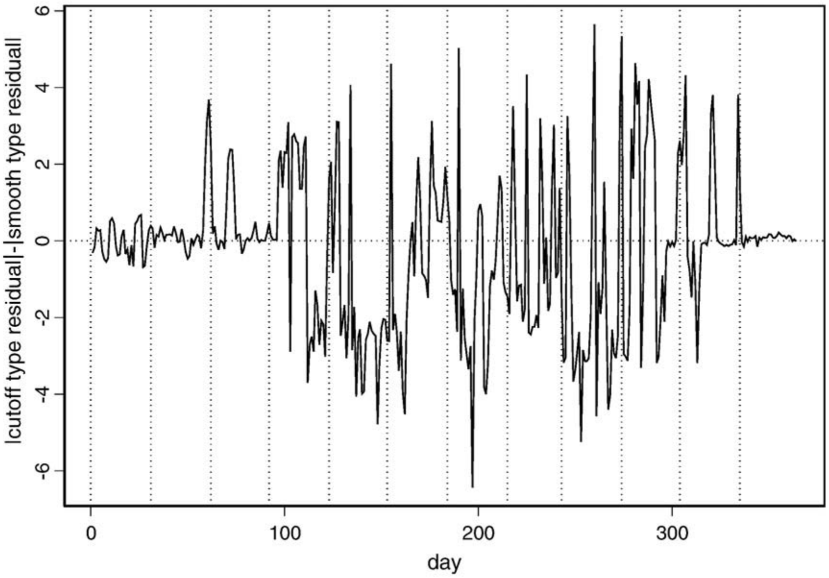


Fig. 10. A comparison of the absolute errors obtained when making predictions of "smooth" and "cutoff" types for an evaluation dataset. Vertical lines denote the beginning of a calendar month, starting with July 2006.

of high consumption. Therefore, we selected $c = 0.75$ as a compromise.

When looking back, one might ask the fundamental question of whether the correction implied by model (2) is needed at all. In other words, would it not suffice to simply fit the conditional model (1) to nonzero data with no corrections whatsoever, and then use it for all predictions of future consumption, for which we do not know whether they are zero or not?

Fig. 11 suggests that the answer is NO. In fact, the full model is better most of the time, and the differences between the two are larger during the summer, as expected. The effect of a simple-minded fitting of the model to days with positive consumption becomes more deleterious during periods when total switch-off is more common.

### 2.1.4. General comments

The approach described in Sections 2.1.1, 2.1.2 and 2.1.3 is a kind of compromise between two extremes that one is typically faced with during routine modeling. The first takes and models each individual separately. The second treats all individuals as if they were really the same, or the same up to a multiplicative constant, with the same average response to temperature and the same weekly consumption pattern. The first is more detailed, but suffers from inefficiency. The second is simple enough, but tends toward over-simplification. NLME is a compromise between the two extremes, in which the weight given to each of them automatically depends on the ratio of inter- and intra-individual variability, or on a kind of analogue to the signal-to-noise ratio.

More specifically, the second approach mentioned above treats all customers as the same, or at least as the same within a particular "segment" of the customer pool. This is far behind the idea of using many standardized load profiles (SLP) (Schwarz et al., 2002; Gerbec et al., 2004) on an individual level. This would amount to using the average instead of individual relationships. Our results show, as a byproduct, that the price to be paid for the simplicity of such an approach is not small, especially in non-residential, inherently heterogeneous segments. The assumption of the same "average dynamics" for everybody, i.e. excluding the interaction between the dynamics' shape and individual effects, neglects a large portion of the overall variability. Thus, the individual and semi-individual approach of NLME provides a substantial improvement.

It is well-known that mixed effects models typically imply the shrinkage of coefficient estimates toward their common population means (Davidian & Giltinan, 1995). This attractive and generally well-perceived

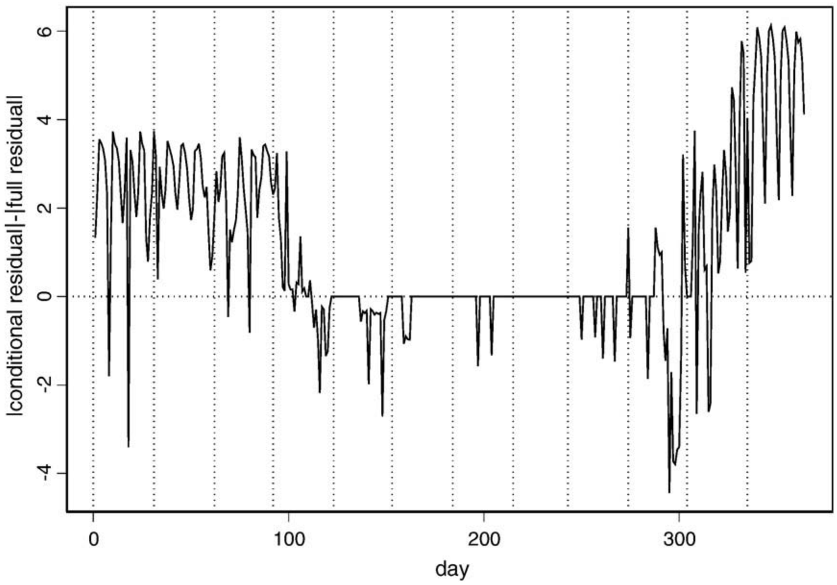


Fig. 11. Comparison of absolute errors obtained from the full model by combining model (1) and model (2) via a "cutoff type" correction $|\sum_i \left(Y_{it} - \hat{Y}_{it,C}\right)|$, and from the conditional model (1) only, with $|\sum_i \left(Y_{it} - g\left(t; \hat{\theta}_i\right)\right)|$ used for the evaluation dataset.

property contributes to the popularity of this model class. Fig. 12 shows the shrinkage of predictions.

It compares the conditional model predictions computed using two methods for two selected customers (A and B). For a particular customer, both predictions are obtained from an Eq. (1)-like model throughout the evaluation period. The first set of predictions was obtained from the NLME model. They are plotted on the vertical axis. Note that the NLME approach uses all data to estimate individual parameters (not only those coming from a particular customer). The second set was obtained from the nonlinear model fitted to each individual customer separately. These predictions are plotted on the horizontal axis. Different time points are plotted as different dots on each plot. The left panel corresponds to a typical customer. It illustrates that the amount of shrinkage is rather small for most customers. There are a few exceptions, and the behavior of one of them is illustrated in the right hand panel. For these exceptions, shrinkage increases progressively with the size of the prediction. Larger predictions are subject to greater shrinkage. From a practical point of view, this is an appealing property. The model does not "believe" in extremes as much as in more typical values. As a result, the NLME predictions are slightly more conservative or robust than the individual fits.

Model (1) implies a nonstationary correlation structure. This is in addition to having a nonstationary and nonlinear mean. Unlike classical TS models, the correlation changes relative to temperature. When temperature decreases, the correlation increases, and vice versa. The implied summer months' correlation decay might be perceived as a realistic feature.

Another substantial attraction of the NLME framework is the inherent exchange ability type assumption. All individual parameters are supposed to come from a common underlying distribution, which has crucial implications for the prediction of new customers. Without this assumption, there would be no reason to expect any similarity between customers having a known history from which their individual parameters can be inferred and new customers for whom no measurements have yet been performed. Therefore, it would be hard to justify any kind of estimate for a newcomer. On the other hand, a population average estimate given as a function of both fixed and random effects arises as a very natural prediction in the NLME context. Moreover, the mixture or hierarchical approach allows for population-level estimates, and thus has an advantage over the piece by piece individual approach of both nonlinear regression and time series models. This might be useful in situations

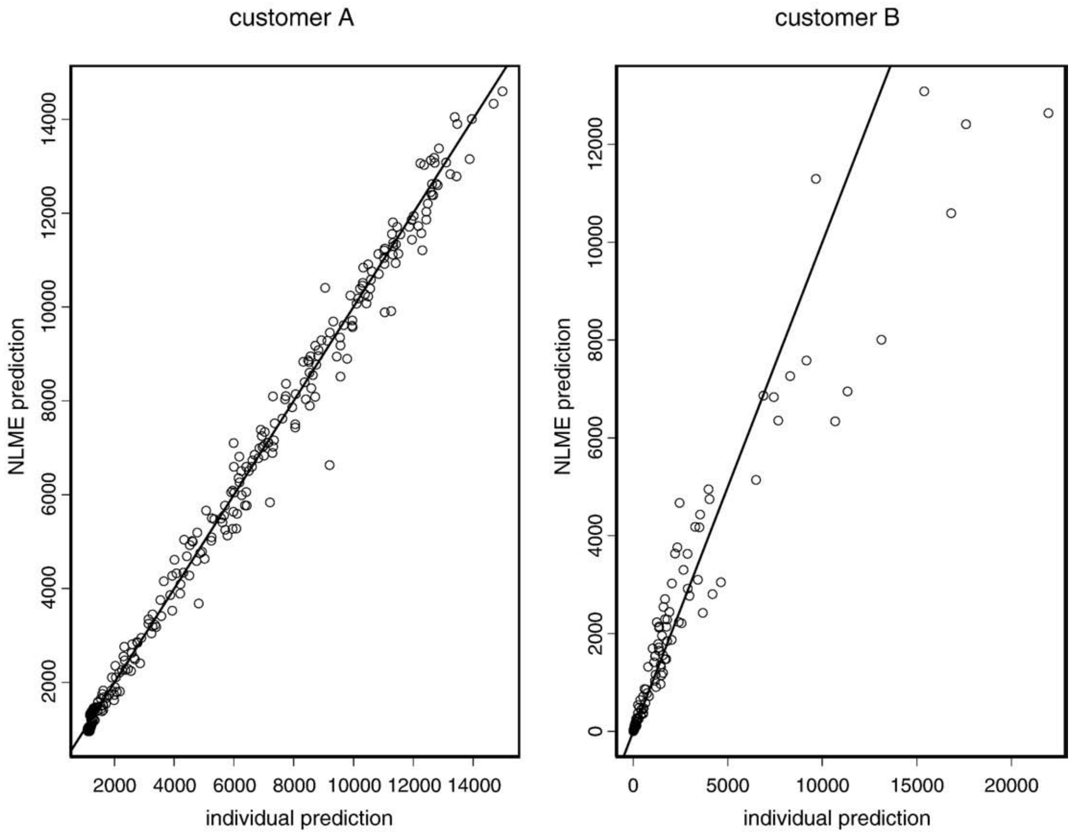


Fig. 12. Shrinkage behavior of model (1) with respect to prediction: the examples of two selected customers. NLME predictions are compared to predictions from an individually fitted model. Each dot corresponds to one time point.

dealing with a sample of individual customers when some inference about the populations they came from is desired. Such computations are hard to justify if one has only a bunch of unrelated individual models. For instance, one might be interested in the estimation of an average consumption trajectory, average trajectory of the $p$-th quantile, etc.

We would like to stress the fact that the population average is a function of both fixed and random effects. This feature is a consequence of the nonlinear character of the model. In particular, $E_i(Y_{it}|Y_{it}>0) \neq g(t;\, E_i(\theta_i))$ holds. The operator $E_i$ denotes the expectation over customers. In other words, $E_i(Y_{it}|Y_{it}>0)$ cannot be obtained simply by evaluating $g(t;.)$ for fixed effects as the parameter means. Averages of nonlinear functions must be evaluated here, e.g., by Monte Carlo simulations or by analytical approximations, such as the Laplace approximation.

Since the mixed model is inherently smooth, both in terms of prediction trajectories and with respect to being influenced by incoming data, one might ask whether it is really necessary to reestimate the model each time a new monthly batch arrives. Is the model “open” to incoming information? In other words, how much would we lose if we estimated the individual-specific parameters only once from the historical data and used these values throughout the evaluation period?

Fig. 13 shows that we would lose a substantial amount of prediction precision, and that re-estimation upon the receipt of fresh data would certainly pay off. See Section 4 for further comments.

### 2.2. Benchmark models

To compare the proposed NLME model with some “standard” approaches, we selected two time series

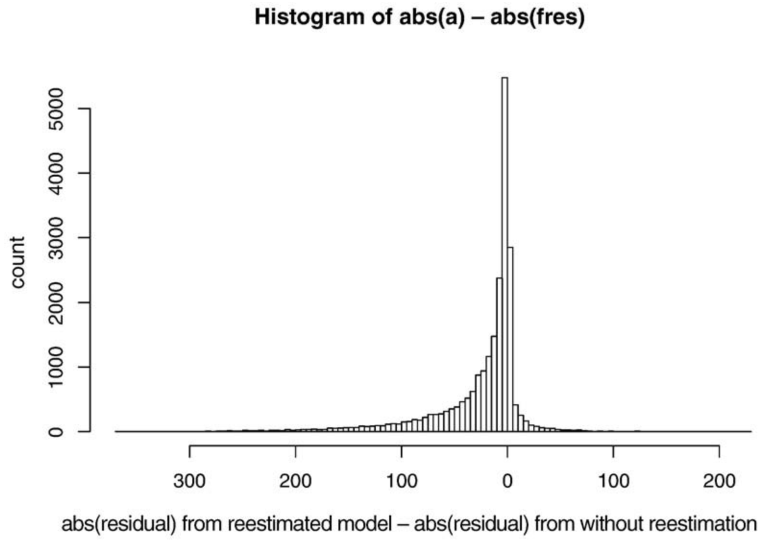


Fig. 13. Histogram of the difference between absolute residuals from reestimated and not reestimated "smooth type" model predictions.

(TS) based models, ARIMAX and ARX, as benchmarks. This choice was motivated by the widespread use and popularity of the Box-Jenkins class of models in practice. The ARIMAX and ARX models were obtained via different model selection strategies. We selected the form of the first by means of automatic model selection within the ARIMAX class, based on AIC. The ARX model was selected "manually" within a more restricted autoregressive class. It was built in a more or less structural fashion, and the selection of its components was based on previous experience with similar data and expert knowledge, as well as common principles used in gas modeling. This is in contrast to the automatic selection approach applied here to the ARIMAX model, for which a wide class was automatically searched to yield a model, in accordance with common practice.

In both benchmark models, the zero readings were considered as valid values. We tested several methods for the replacement of missing values with practically identical results. We decided to replace missing observations by their predictions.

### 2.2.1. ARIMAX

We modeled the consumption $Y_{it}$ of customer $i$ on day $t$ using an ARIMA model with exogenous variables. The order of the ARIMA model was optimized by the minimization of the AIC using the means of all customers' daily consumption as a learning dataset. Similarly to the NLME case, we used the average daily temperature as an explanatory variable. Consistent with extensive local experience, historical data behavior (see Fig. 7), and our previously published results (Vondráček et al., 2008), the temperatures $T_t$ on day $t$ were cut at the level of 14 °C as follows:

$$\widetilde{T}_t = \begin{cases} T_t & \text{if } T_t < 14 \\ 14 & \text{if } T_t \geq 14. \end{cases} \tag{3}$$

An ARIMA(3, 1, 3)(0, 1, 1)$_7$ model with exogenous variables (average daily temperature $\widetilde{T}_t$ on day $t$ and lagged temperature $\widetilde{T}_{t-1}$) was then fitted to every individual customer's consumption time series. The model has the form:

$$Y_{it} = b_{i1}\widetilde{T}_t + b_{i2}\widetilde{T}_{t-1} + R_{it}. \tag{4}$$

The residuals $R_{it}$ follow the ARIMA(3, 1, 3)(0, 1, 1)$_7$ model according to the formula:

$$\begin{aligned}&\left(1 - a_{i1}B - a_{i2}B^2 - a_{i3}B^3\right)(1-B)\left(1-B^7\right)R_{it} \\ &\quad = \left(1 - m_{i1}B - m_{i2}B^2 - m_{i3}B^3\right)\left(1 - M_{i1}B^7\right)e_{it},\end{aligned} \tag{5}$$

where $a_{i1}, a_{i2}, a_{i3}, m_{i1}, m_{i2}, m_{i3}$, and $M_{i1}$ are non-random model parameters, $B$ is the backward shift operator, e.g. $BR_{it}=R_{i,t-1}$, and $e_{it}$ is normal white noise. The parameters were optimized via the arimax function of the R package TSA (Cryer & Chan, 2008).

*2.2.2. Linear ARX model*

For each individual customer $i$, our ARX model, i.e. the autoregressive model with external inputs (Box & Jenkins, 1990), has the form:

$$Y_{it} = a_{i1} Y_{i,\,t-1} + a_{i2} Y_{i,\,t-2}, + a_{i7} Y_{i,\,t-7} + b_{i0} \widetilde{T}_t + b_{i1} \widetilde{T}_{t-1} + c_{i0} + e_{it} \tag{6}$$

where $a_{i1}$, $a_{i2}$, $a_{i7}$, $b_{i0}$, $b_{i1}$, and $c_{i0}$ are non-random model parameters, and $e_{it}$ is the system noise, which is supposed to have a normal distribution with a mean of zero and a constant variance. The relationship between the individual consumption $Y_{it}$ and the actual daily averaged temperature $\tilde{T}_t$, the temperature of the previous day $\tilde{T}_{t\text{-}1}$, and the past consumption values was modeled by linear regression. Temperatures were cut according to the formula (3). The explanatory variables were selected by analyzing the total consumption time series, i.e., the sum of all individual consumptions. Here, we applied the statistical variable selection procedure implemented in the FORECAST PRO software (Business Forecast Systems, Inc., 2008). This results in a model form that is the same for all individuals, but its parameters are individual-specific. This was done deliberately in order to be able to later compare NLME, ARX and ARIMAX models based on the same (individual) grounds.

*2.2.3. General comments about the benchmark models*

Note that the ARX and ARIMAX models are of different types. The first is stationary, except for the external explanatory variables part. The latter is in differenced, i.e. nonstationary, form. NLME is non-stationary as well.

TS models can be expanded by conditioning on the non-zero consumption status, similar to the approach utilized in Section 2.1. To this end, we tried to fit the TS models to the non-zero data only, and then combined it with the “cutoff” type of zero consumption probability portion, similar to Section 2.1.3. This modification, however, did not lead to any substantial improvement in terms of individual MAE.

## 3. Prediction performance

In this section, we compare the prediction performance of the four models described in Section 2, based on an evaluation dataset. We will call the “smooth type” of the NLME model “NLME smooth” and the

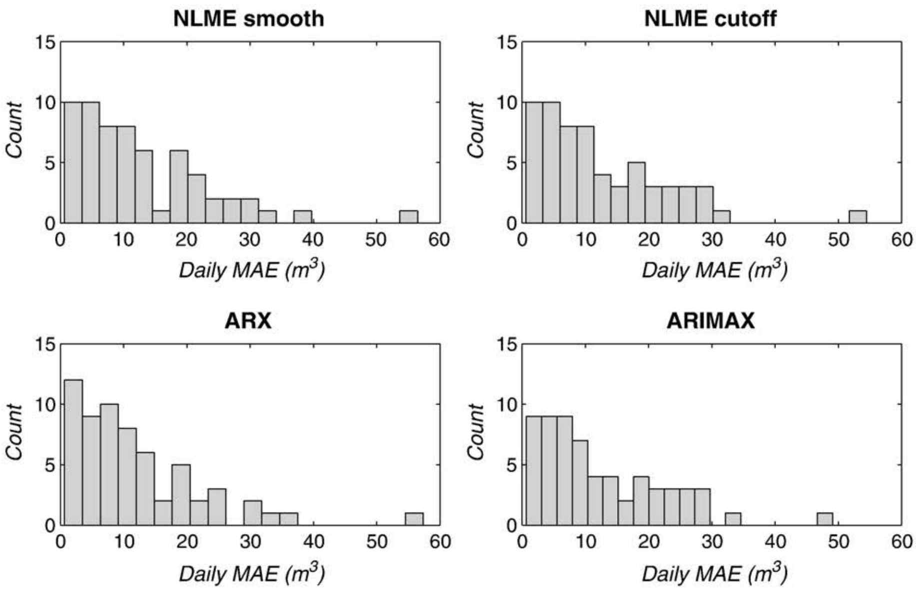


Fig. 14. Histograms of individual MAEs — a comparison of our four models.

Table 1
The averages of individual MAEs across all customers for the four models (in $m^3$)

| NLME "smooth" | NLME "cutoff" | ARX | ARIMAX |
|---|---|---|---|
| 12.73 | 12.60 | 11.95 | 12.49 |

"cutoff type" of the NLME model "NLME cutoff"; they are described in Sections 2.1 and 2.1.3, respectively. The "ARIMAX" ARIMA model with exogenous variables is described in Section 2.2.1, and "ARX", the ARX model, is described in Section 2.2.2.

### 3.1. Individual prediction

Daily post-sample forecast errors, i.e. errors for each day of the evaluation period, with varying forecast horizons, were computed for each model and each customer. Histograms of the individual mean absolute errors

$$MAE_i = \frac{\sum_{t\in T_i} |\hat{Y}_{it} - Y_{it}|}{|T_i|}, \tag{7}$$

where $T_i$ is the set of non-missing data from the evaluation period for customer $i$, and $\hat{Y}_{it}$ is the forecast for consumption $Y_{it}$ from the particular model, are shown in Fig. 14.

In Table 1, we compare the MAEs (averages of the $MAE_i$ across all individuals) of the four models. We can see that the MAEs are quite similar, and that, on average, the forecasting performances of the models are comparable.

To check that the models did not grossly over-estimate or underestimate in a systematic way, and to compare them from this viewpoint, we plotted histograms of all individual prediction errors ($\hat{Y}_{it}$–$Y_{it}$) in Fig. 15. For all four models, the majority of the individual daily errors are close to zero and the histograms are relatively symmetric. From a practical point of view, there seems to be no substantial difference among the models.

### 3.2. Performance with respect to the forecast horizon

To test the model performance with respect to the forecast horizon, we evaluated the post-sample forecast errors for each model, each customer, and each day. The mean absolute error (MAE) for the corresponding forecast horizon was then computed. Averaging was performed across days having the same forecast horizon

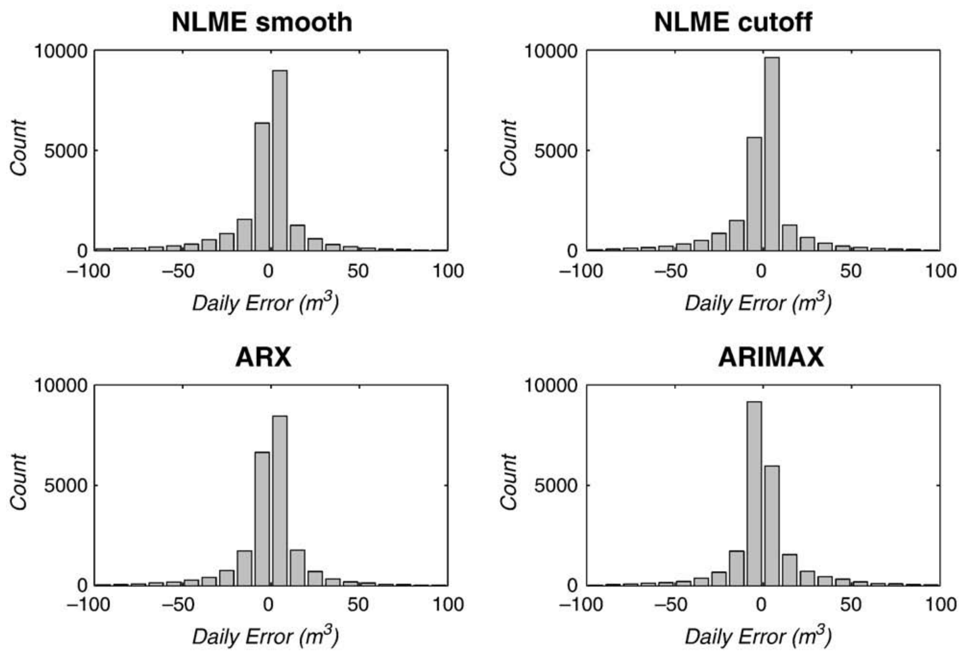


Fig. 15. Histograms of all daily prediction errors — a comparison of our four models.

(i.e. same order in the calendar month) and across all customers. The horizon varied from 1 to 31 days, according to the current calendar month. The resulting plots of MAE vs. forecast horizon can be seen in Fig. 16. Table 2 shows the values of the MAEs for the four models, with a breakdown based on forecast horizon. It is apparent that the performance is quite consistant for the NLME and ARX approaches, in contrast to the ARIMAX approach, for which the MAE grows with the forecast horizon, even with the use of exogenous variable information. Both TS models perform well for very short prediction horizons. The fact that the ARIMAX approach shows some instability at longer prediction horizons might be tied to the fact that the model form was selected automatically and forced to be the same for all individuals (in order to be compatible with NLME, where only the parameters are individual-specific, not the model form). This finding can be taken as a warning against relying too much on the automatic selection. Individually, some difficulties might occur with MA and AR polynomial factor near-cancellation, but in general the problem is quite local in time. It is related to the ability of the model to capture the transient phenomenon of the spring consumption decline, and, to a much smaller extent, the autumn consumption increase, as can be seen in Fig. 17. In addition, one might suspect that the quality of the temperature as a predictor is much lower during these periods, due to the very heterogeneous temperature response of individual customers.

### 3.3. Performance with respect to the calendar month

We also tested the prediction performance in terms of the calendar month. In Fig. 17 and Table 3, we compared our four models using the daily MAE, i.e. the averages of forecast errors within the corresponding month. The resultant averaging across lead times is, in general, not very amenable to the generic forecasting performance. However, in our situation averaging is implied by the practical problem at hand. This is because one deals with different horizons due to the fact that the individual days of the next month need to be predicted from the same origin — the last day of the current month — before new data are available, as was described in Section 1. Therefore, averaging across different days is natural here, and corresponds to the "overall performance" of the forecasting procedure with respect to the problem.

The NLME approach is preferable to time series approaches in Summer (June to August), and slightly preferable in Spring (namely April). In Winter (December to February), the time series approaches have lower MAEs. MAEs are related to the consumption level, and are lower in summer and higher in winter for all four models.

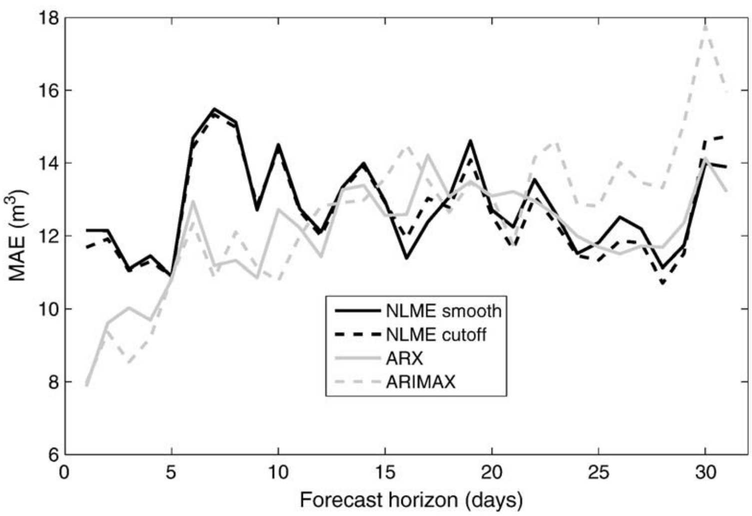


Fig. 16. Daily MAE vs. the forecast horizon — a comparison of our four models.

## 4. Conclusion

Our proposed NLME model stresses individuality smoothly. It makes an automatic compromise between all-the-same and fully individual dynamics by shrinking. This is more advantageous, and indeed more appealing, the greater the similarity among individuals. On the other hand, individual time series approaches do not allow for shrinking, and hence might be a bit too variable. Nevertheless, they are good for local adaptation to sudden and abrupt changes within a single individual. This is something that the NLME just cannot do.

We observed that the NLME approach and individual time series (ARIMAX, ARX) approaches all have their merits and their problems. In fact, when we did some evaluation using the Diebold-Mariano test (Diebold & Mariano, 1995) to compare them, there was no clear winner. One should realize that the comparisons performed here should not be generalized to all time series models or to all NLME-based models. In particular, we stress the fact that the two TS models were selected as examples mimicking procedures likely to be used in practice (based on simplified model selection) that might not be completely optimal.

Table 2
MAE in $m^3$ vs. forecast horizon

| Horizon [days] | NLME "smooth" | NLME "cutoff" | ARX | ARIMAX |
|---|---|---|---|---|
| 1 | 12.15 | 11.68 | 7.87 | 7.96 |
| 2 | 12.14 | 11.91 | 9.61 | 9.36 |
| 3 | 11.09 | 11.05 | 10.03 | 8.54 |
| 4 | 11.45 | 11.29 | 9.69 | 9.20 |
| 5 | 10.89 | 10.91 | 10.81 | 10.87 |
| 6 | 14.68 | 14.43 | 12.93 | 12.33 |
| 7 | 15.48 | 15.32 | 11.18 | 10.85 |
| 8 | 15.12 | 14.98 | 11.32 | 12.10 |
| 9 | 12.71 | 12.83 | 10.85 | 11.13 |
| 10 | 14.50 | 14.38 | 12.71 | 10.77 |
| 11 | 12.74 | 12.68 | 12.22 | 12.00 |
| 12 | 12.11 | 11.99 | 11.42 | 12.82 |
| 13 | 13.34 | 13.25 | 13.27 | 12.91 |
| 14 | 13.99 | 13.91 | 13.38 | 12.98 |
| 15 | 12.94 | 12.90 | 12.56 | 13.57 |
| 16 | 11.38 | 11.93 | 12.58 | 14.50 |
| 17 | 12.38 | 13.03 | 14.22 | 13.52 |
| 18 | 13.06 | 12.78 | 13.08 | 12.64 |
| 19 | 14.61 | 14.09 | 13.49 | 13.46 |
| 20 | 12.71 | 12.55 | 13.09 | 13.07 |
| 21 | 12.24 | 11.63 | 13.21 | 11.75 |
| 22 | 13.54 | 13.08 | 12.96 | 14.14 |
| 23 | 12.61 | 12.32 | 12.57 | 14.62 |
| 24 | 11.51 | 11.45 | 11.99 | 12.87 |
| 25 | 11.81 | 11.33 | 11.69 | 12.81 |
| 26 | 12.51 | 11.86 | 11.50 | 14.01 |
| 27 | 12.19 | 11.80 | 11.72 | 13.46 |
| 28 | 11.13 | 10.71 | 11.68 | 13.32 |
| 29 | 11.74 | 11.51 | 12.35 | 15.09 |
| 30 | 13.99 | 14.62 | 14.13 | 17.76 |
| 31 | 13.89 | 14.73 | 13.21 | 15.95 |

Time series modeling based predictors tend to be less smooth, or more variable, over time. On the other hand, this is not true for prediction errors. For instance, if we think about $\hat{E}(Y_{it}|\mathcal{Y}_{i,t-1})$, ARIMAX tends to have σ-algebra of past events ($\mathcal{Y}_{i,t-1}$) that is larger than NLME, at least in our case. This is clearly apparent when we compare the standard deviations of predictions averaged across customers, i.e. $s_{\hat{Y}_i} = \sqrt{\sum_{i=1}^{n} s_i^2/n}$, where $s_i$ is the standard deviation of $\hat{Y}_i$s over the twelve months of the evaluation period. For the "smooth type" mixed model (Section 2.1.3), we get $s_{\hat{Y}_i}=41.6\ m^3$, while for ARIMAX (Section 2.2.1) we get $s_{\hat{Y}_i}=47.6\ m^3$. This might be advantageous in some cases, e.g. when the weekly consumption pattern changes frequently within the same individual. On the other hand, the smoothness built into the NLME modeling, as well as the "borrowing of strength" across individuals that it implies, will be beneficial in situations when there is more regularity in the intra- and/or inter-individual sense.

Ideally, one might want to somehow combine the two approaches to include both the shrinking feature and local flexibility. The first appears to be due to the presence of random effects, while the latter is due to a time series structure that is present in the model. Such a combination could lead, for example, to a nonlinear mixed effects model, with a nontrivial time series structure in the residuals. We would like to consider this in future work.

Estimation and prediction in the combined model is nontrivial, mainly from a computational point of view. Thus far we have tried to expand the conditional model (1) by introducing the AR(1) process into its residuals, and faced substantial convergence problems that probably would not appear in cases without residual autocorrelation. Another possibility would be to have a time-series structure in the parameters to capture their changes over time. We plan to approach this via the state-space approach and an extended Kalman filter. In

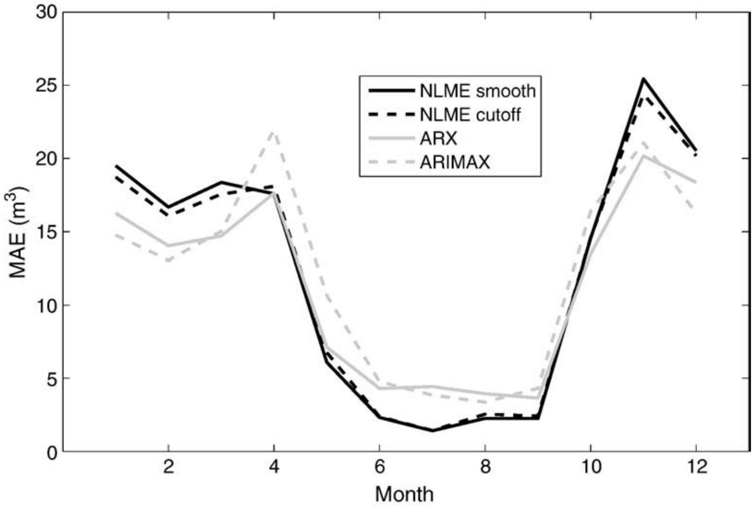


Fig. 17. Daily MAE vs. calendar month — a comparison of our four models.

particular, one might let a generic parameter of model (1), say $\theta$, be time-dependent and have an individual state equation. The random walk would be the most simple possibility to consider here: $\theta_{it}=\theta_{i,t-1}+\mathrm{error}_{it}$. Nevertheless, it remains to be seen whether the additional complexity provoked by the more comprehensive and elegant model formulation provides sufficiently large improvements to the predictions to be truly of practical value over the models and methods used in this paper.

Another popular approach that might, in principle, be used in a situation such as this is to combine different estimators or predictors. However, we prefer to first improve each model class separately as much as possible. The resulting models might then be combined if it is necessary and would be advantageous. We consider the path leading through individual model improvement, especially in the case of a structured model with interpretable components like our NLME model, to be a bit cleaner and clearer. We also view it as more "controllable", because adding, changing, or removing the structural blocks has predictable effects. On the other hand, combining totally different models based on historical calibration might be very good for some situations, but unsafe for others. Based on this, we consider the structural improvement path to be safer than the other, more pragmatic, but obviously often very successful, approach.

Table 3
MAE in $m^3$ vs. month

| Month | NLME "smooth" | NLME "cutoff" | ARX | ARIMAX |
|---|---|---|---|---|
| 1 | 19.5074 | 18.7263 | 16.2564 | 14.7639 |
| 2 | 16.6771 | 16.0698 | 14.0282 | 13.0491 |
| 3 | 18.3603 | 17.5574 | 14.7029 | 15.0197 |
| 4 | 17.5845 | 18.0888 | 17.6218 | 21.9709 |
| 5 | 6.0909 | 6.7702 | 7.1238 | 10.6531 |
| 6 | 2.3217 | 2.3716 | 4.3045 | 4.7615 |
| 7 | 1.4112 | 1.4322 | 4.4337 | 3.8451 |
| 8 | 2.2590 | 2.5483 | 3.9462 | 3.3644 |
| 9 | 2.2406 | 2.4137 | 3.6514 | 4.3075 |
| 10 | 14.5382 | 14.5953 | 13.4740 | 16.4011 |
| 11 | 25.4380 | 24.3885 | 20.1631 | 21.0616 |
| 12 | 20.5246 | 20.1891 | 18.3557 | 16.2453 |

## Acknowledgements

This paper was supported by GA AS CR grant No. 1ET400300513 and by the institutional research plan of ICS AS CR No. AV0Z10300504.

The authors wish to thank the Slovak Gas Industry company (namely Ing. Rastislav Prelec) for valuable discussion and for introducing us to the practical problem. We would like to thank two anonymous referees and the editor of this special issue, Antoni Espasa, for valuable comments and important insights that helped us to substantially improve the original manuscript version.

## References

Amaral, L. F., Souza, R. C., & Stevenson, M. (2008). A smooth transition periodic autoregressive (STPAR) model for short-term load forecasting. *International Journal of Forecasting*, *24*, 603−615 (this issue).

Anonymous (2001). *S-Plus 6 for Windows, Guide for Statisticians (Volume I).* Seattle, Washington: Insightful.

Bailey, J. (2000). Load Profiling for Retail Choice: Examining a Complex and Crucial Component of Settlement. *The Electricity Journal*, *13*, 69−74.

Bartels, R., Fiebig, D. G., & Nahm, D. (1996). Regional end-use gas demand in Australia. *Economic Record*, *72*, 319−331.

Bauwens, L., Fiebig, D. G., & Steel, M. F. J. (1994). Estimating end-use demand: A Bayesian approach. *Journal of Business and Economic Statistics*, *12*, 221−231.

Box, G. E. P., & Jenkins, G. (1990). *Time series analysis, forecasting and control.* San Francisco: Holden-Day.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, *88*, 9−25.

Business Forecast Systems, Inc. (2008). *Forecast Pro.* Business Forecast Systems, Inc., Belmont, MA. http://www.forecastpro.com. Accessed on 30 April 2008.

Cancelo, J., Espasa, A., & Grafe, R. (2008). Forecasting from one day to one week ahead for the Spanish system operator. *International Journal of Forecasting*, *24*, 588−602 (this issue).

Cottet, R., & Smith, M. (2003). Bayesian modeling and forecasting of intraday electricity load. *Journal of the American Statistical Association*, *98*, 839−849.

Cryer, J. D., & Chan, K. S. (2008). *Time series analysis with applications in R.* New York: Springer.

da Silva, A. P. A., Ferreira, V. H., & Velasquez, R. M. G. (2008). Input space to neural network based load forecasters. *International Journal of Forecasting*, *24*, 616−629 (this issue).

Davidian, M., & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data.* London: Chapman and Hall.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*, 253−263.

Dordonnat, V., Koopman, S. J., Ooms, M., Dessertaine, A., & Collet, J. (2008). An hourly periodic state space model for modelling French national electricity load. *International Journal of Forecasting*, *24*, 566−587 (this issue).

Gerbec, D., Gašperič, S., Šmon, I., & Gubina, F. (2004). Determining the load profiles of consumers based on fuzzy logic and probability neural networks. *IEEE Proceedings — Generation, Transmission and Distribution*, *151*, 395−400.

Ghosh, J. K., Delampady, M., & Samanta, T. (2006). *An introduction to Bayesian analysis.* New York: Springer.

Gil, S., & Deferrari, J. (2004). Generalized model of prediction of natural gas consumption. *Transactions of the ASME*, *126*, 90−97.

Harvey, A., & Koopman, S. J. (1993). Forecasting hourly electricity demand using time-varying splines. *Journal of the American Statistical Association*, *88*, 1228−1236.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: John Wiley & Sons.

Liu, J. M., Chen, R., Liu, J. M., Liu, L. M., & Harris, J. L. (2006). A semi-parametric time-series approach in modeling hourly electricity loads. *Journal of Forecasting*, *25*, 537−559.

McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models.* New York: John Wiley.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-plus.* New York: Springer.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.* http://www.r-project.org/. Accessed on 30 April 2008.

Sánchez-Úbeda, E. F., & Berzosa, A. (2007). Modeling and forecasting industrial end-use natural gas consumption. *Energy Economics*, *29*, 710−742.

Schwarz, D., Kubessa, M., & Fuhrberg-Baumann, J. (2002). Load profiles and regression functions for customers not measured according to output in the liberalized gas market.*Gas- und Wasserfach*, *143*, 689−692 (in German).

Searle, S. R. (1971). *Linear models.* New York: John Wiley.

Soares, L. J., & Medeiros, M. C. (2008). Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data. *International Journal of Forecasting*, *24*, 630−644 (this issue).

Suykens, J., Lemmerling, P., Favoreel, W., De Moor, B., Crepel, M., & Briol, P. (1996). Modelling the Belgian gas consumption using neural networks. *Neural Processing Letters*, *4*, 157−166.

Taylor, J. W. (2008). An evaluation of methods for very short-term load forecasting using minute-by-minute British data. *International Journal of Forecasting*, *24*, 645−658 (this issue).

Venables, W. N., & Ripley, B. D. (1999). *Modern applied statistics with S-PLUS (3rd ed.).* New York: Springer.

Vondráček, J., Pelikán, E., Konár, O., Čermáková, J., Eben, K., Malý, M., et al. (2008). A statistical model for the estimation of natural gas consumption. *Applied Energy*, *85*, 362−370.

Wang, G. C. S. (2004). Forecasting practices in electric and gas utility companies. *Journal of Business Forecasting Methods and Systems*, *23*, 11−15.

Wolfinger, R., & O'Connell, M. (1993). Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, *48*, 233−243.